

# REVUE DE STATISTIQUE APPLIQUÉE

T. MOREAU

J. LELLOUCH

## **Données de survie avec variables pronostiques : analyse en présence de nombreux ex-æquos parmi les temps de survie**

*Revue de statistique appliquée*, tome 29, n° 2 (1981), p. 5-14

[http://www.numdam.org/item?id=RSA\\_1981\\_\\_29\\_2\\_5\\_0](http://www.numdam.org/item?id=RSA_1981__29_2_5_0)

© Société française de statistique, 1981, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

# DONNEES DE SURVIE AVEC VARIABLES PRONOSTIQUES : ANALYSE EN PRESENCE DE NOMBREUX EX-AEQUOS PARMI LES TEMPS DE SURVIE

T. MOREAU et J. LELLOUCH

INSERM U.169, 16bis, Avenue Paul Vaillant Couturier, 94800 – Villejuif

## RESUME

Les principales méthodes d'analyse des données de survie éventuellement censurées avec prise en compte de variables pronostiques sont rappelées dans leur principe. Seules sont considérées ici les méthodes ne supposant pas un modèle paramétrique simple (exponentiel, Weibull...) pour les fonctions de survie. Lorsqu'un grand nombre de sujets sont morts à un même temps, la plupart de ces méthodes ne peuvent être appliquées, en raison de temps de calcul trop élevés. Deux méthodes sont adaptées à ce type de données : l'une est due à PRENTICE et GLOECKLER [10], l'autre à THOMPSON Jr. [11]. Une simplification de cette dernière méthode est développée ici ; appliquée au cas particulier de la comparaison de deux groupes, elle nous a conduit à la construction d'une statistique de test très voisine de celle proposée par MANTEL [9]. Ces différentes méthodes sont ensuite comparées sur un exemple numérique ; les résultats trouvés apparaissent très voisins.

## 1. INTRODUCTION

Les études de survie concernant un groupe de sujets sont généralement organisées de la façon suivante : chaque sujet entre dans l'étude au fur et à mesure que celle-ci se déroule et est surveillé pendant une certaine durée à l'issue de laquelle ce sujet est soit mort, soit encore vivant. Dans le premier cas, la durée de surveillance est appelée temps de survie, dans le deuxième cas temps de censure, le sujet correspondant étant dit censuré ; le temps de survie exact d'un sujet censuré est ainsi inconnu, la seule information disponible est qu'il est supérieur au temps de censure. Dans tout ce qui suit, le mécanisme de censure est supposé indépendant des temps de survie, et l'évènement "c est un temps de censure" est identique à l'évènement ; " $T > c$ ", où T est la variable aléatoire temps de survie.

Le problème d'évaluer l'influence de certains facteurs de risque (appelé par la suite covariables) sur les fonctions de survie des sujets surveillés est central à de telles études.

Plusieurs méthodes générales permettant de résoudre ce problème, ont été décrites. Elles sont rappelées dans le § 2.

Le but de cet article est de décrire des méthodes utilisables en présence de nombreux ex-aequos parmi les temps de survie. Il sera vu au § 3, que, dans

ce cas, les méthodes ci-dessus ne peuvent pas être appliquées, pour des raisons soit théoriques soit pratiques. Deux autres méthodes, décrites au §3 peuvent être utilisées dans cette situation : la première est due à PRENTICE et GLOECKLER [10] et la deuxième à THOMPSON Jr. [11]. Nous proposons ici d'utiliser une simplification de la méthode de THOMPSON Jr.

La comparaison des différentes méthodes est effectuée au §4 sur un exemple utilisant des données réelles. Les résultats obtenus apparaissent très voisins.

## 2. METHODES GENERALES

Formellement, les valeurs des covariables mesurées sur le sujet "i" peuvent s'écrire sous forme de vecteur :

$$z'_i = (z_{i1}, z_{i2}, \dots, z_{ip})$$

où  $p$  est le nombre de covariables considérées ; soit  $T_{z_i}$  la variable aléatoire : "temps de survie du sujet i",  $(S(t, z_i))$  la fonction survie ( $S(t, z_i) = \Pr [T_{z_i} > t]$ ) et  $\lambda(t, z_i)$  la fonction "risque instantané" (r.i.) définie, lorsque  $T_{z_i}$  est continue, par :

$$\lambda(t, z_i) = \frac{-dS(t, z_i)}{S(t, z_i)}$$

Lorsque  $T_{z_i}$  est discrète, prenant ses valeurs en  $t_1, t_2, \dots, t_k$ , les quantités suivantes sont définies :

$$p_j(z_i) = P(T_{z_i} = t_j | T_{z_i} > t_{j-1}), \quad j = 1, 2, \dots, k$$

L'approche non paramétrique considérée ici mène à poser que la relation liant  $\lambda(t, z_i)$  (resp.  $p_j(z_i)$ ) à  $t$  (resp. à  $t_j$ ) n'est pas connue, tandis que celle liant  $\lambda(t, z_i)$  (resp.  $p_j(z_i)$ ) à  $z_i$  est analytiquement connue. C'est ainsi que COX [5], dans un important article, a proposé, quand  $T_{z_i}$  est continue, la relation suivante :

$$\lambda(t, z_i) = \exp(\beta' z_i) \times \lambda_0(t) \quad (1)$$

où  $\beta' = (\beta_1, \beta_2, \dots, \beta_p)$  est un vecteur de paramètres inconnus, et  $\lambda_0(t)$  n'est pas précisée.

Il apparaît dans (1) que chaque variable agit "exponentiellement" sur la fonction r.i. ; cette hypothèse, qui peut être testée (voir p. ex. [8]) généralise l'hypothèse dite des "proportionnal hazards" [4] utilisée pour la comparaison des survies de plusieurs groupes.

Le modèle (1) sert de base aux deux méthodes décrites au § 2.1.

### 2.1. Méthodes utilisant un modèle continu

Dans ce paragraphe, les variables aléatoires  $T_{z_i}$  sont supposées continues.

Les temps de survie observés sont désignés par  $t_1, t_2, \dots, t_k$ .

## Méthode I

Cette méthode s'applique lorsqu'un seul sujet meurt à chaque  $t_j$ ,  $j = 1, \dots, k$  (cas sans ex-aequos). En se plaçant aux  $t_j$  fixés, COX [5], [6] utilise une vraisemblance partielle calculée comme suit : soit  $R_j$  l'ensemble des sujets "à risque" à  $t_j$ , c'est-à-dire l'ensemble des sujets ni censurés ni morts juste avant  $t_j$ . Alors conditionnellement au fait que le sujet "j" appartient à  $R_j$ , la probabilité que son temps de survie soit compris dans l'intervalle  $t_j, t_j + dt$  est donné par :

$$\lambda(t_j, z_j) \times dt = \lambda_0(t_j) \times \exp(\beta' z_j) dt$$

par suite, la probabilité de mort en  $t_j$  du sujet "j", connaissant tous les sujets à risque à  $t_j$ , c'est-à-dire à  $R_j$  fixé, vaut sous le modèle (1) :

$$\frac{\lambda_0(t_j) \times \exp(\beta' z_j) dt}{\sum_{i \in R_j} \lambda_0(t_j) \exp(\beta' z_i) dt} = \frac{\exp(\beta' z_j)}{\sum_{i \in R_j} \exp(\beta' z_i)}$$

Le produit de tels termes pour  $j = 1, 2, \dots, k$  fournit une vraisemblance dont le Log est :

$$L_1(\beta) = \sum_{j=1}^k \beta' z_j - \sum_{j=1}^k \text{Log} \left\{ \sum_{i \in R_j} \exp(\beta' z_i) \right\}$$

qui ne dépend pas de  $\lambda_0(t)$ ;  $L_1(\beta)$  ayant les mêmes propriétés asymptotiques qu'une vraisemblance ordinaire, il est possible d'obtenir numériquement un estimateur  $\hat{\beta}$  de  $\beta$  maximisant  $L_1(\beta)$ ; les calculs sont effectués au moyen d'une méthode itérative, comme par exemple la méthode Newton-Raphson.

D'autre part, en notant  $W(\beta)$  la matrice dont l'élément  $(r, s)$  vaut :  $(-\partial^2 L_1 / \partial \beta_r \partial \beta_s)$ , et par  $V(\beta)$  le vecteur-colonne des dérivées premières de  $L_1(\beta)$  par rapport aux  $\beta_i$ ,  $i = 1, \dots, p$ , un test de  $H_0 : (\beta = 0)$  est donné par la statistique :

$$V(0)' W(0) V(0) \quad (2)$$

dont la distribution asymptotique est un  $\chi_p^2$  sous  $H_0$ .

## Méthode II

Cette méthode, due à BRESLOW [3], repose sur le modèle (1) avec la condition :

$$\lambda_0(t) = \lambda_j \text{ pour } t_{j-1} < t \leq t_j \text{ avec } t_0 = 0$$

Bien que cette méthode soit ainsi paramétrique, la condition ci-dessus est d'autant moins restrictive sur la forme des fonctions de survie, que les intervalles  $(t_{j-1}, t_j)$  sont petits, c'est pourquoi la méthode est décrite ici.

BRESLOW calcule la vraisemblance ordinaire, à laquelle un sujet censuré entre  $t_j$  et  $t_{j+1}$  contribue par sa fonction de survie en  $t_j$ ; si  $l_2(\beta, \lambda_1, \lambda_2, \dots, \lambda_k)$  désigne le Log de cette vraisemblance, la quantité :

$$L_2(\beta) = \sup_{\lambda_1, \lambda_2, \dots, \lambda_k} l_2(\beta, \lambda_1, \lambda_2, \dots, \lambda_k)$$

fournit une vraisemblance pour  $\beta$  ; le calcul montre que :

$$L_2(\beta) = \sum_{j=1}^k \left[ \beta' S_j - m_j \text{Log} \left\{ \sum_{i \in R_j} \exp(\beta' z_i) \right\} \right]$$

où  $m_j$  est le nombre de sujets morts en  $t_j$ , et  $S_j$  est un vecteur qui est la somme des vecteurs de covariables correspondant aux sujets morts en  $t_j$ .

Un estimateur  $\hat{\beta}$  de  $\beta$  maximisant  $L_2(\beta)$  peut être calculé de la même façon que dans la Méthode I.

## 2.2. Méthodes utilisant un modèle discret

### Méthode III

Afin de rendre compte de la présence, fréquente en pratique, d'ex-aequo parmi les temps de survie, COX [5] a proposé un modèle où les variables aléatoires  $T_{z_i}$  sont discrètes, prenant leur valeur aux temps de survie observés  $t_j$ ,  $j = 1, \dots, k$ . Ce modèle s'écrit, avec les notations du § 2 :

$$\frac{p_j(z_i)}{1 - p_j(z_i)} = \exp(\beta' z_i) \frac{p_j}{1 - p_j} \quad (3)$$

où  $p_j = p_j(0)$ .

L'équation (3) peut être considérée comme une approximation de l'équation (1). L'auteur suggère l'utilisation de la vraisemblance partielle, comme suit : soient  $n_j$  le nombre de sujets à risque à  $t_j$ ,  $M_j$  l'ensemble des  $m_j$  sujets morts à  $t_j$  et  $S_j$ , comme plus haut, le vecteur somme des vecteurs de covariables correspondant à ces sujets.

Soit aussi  $S_{lj}$  le vecteur somme des vecteurs de covariables correspondant aux  $m_j$  sujets de l'un des sous ensembles de  $m_j$  sujets pris parmi les  $n_j$  sujets de  $R_j$  ;  $l$  varie de 1 à  $n_j! / (n_j - m_j)! m_j!$ .

Alors, conditionnellement à  $R_j$  et  $m_j$ , la probabilité que les sujets de  $M_j$  meurent à  $t_j$  est :

$$\frac{\exp(\beta' S_j)}{\sum_1 \exp(\beta' S_{lj})}$$

Comme dans la méthode I, le produit de tels termes pour  $j = 1, 2, \dots, k$  fournit une vraisemblance, dont le Log s'écrit :

$$L_3(\beta) = \sum_{j=1}^k \beta' S_j - \sum_{j=1}^k \text{Log} \left\{ \sum_1 \exp(\beta' S_{lj}) \right\}$$

et  $L_3(\beta)$  ne dépend pas des  $p_j$ .

Il est possible de calculer un estimateur  $\hat{\beta}$  de  $\beta$  et de former un test de  $H_0 : (\beta = 0)$  de la même façon qu'en 2.1.

Dans le cas simple de la comparaison de deux groupes, où  $z_i = 0$  ou 1 selon que le sujet "i" appartient à un groupe ou à l'autre, le test obtenu est identique au test de MANTEL [9].

#### Méthode IV

KALBFLEISCH et PRENTICE [7] ont considéré le cas où les variables aléatoires  $T_{z_i}$  sont continues satisfaisant (1), et où les données sont recueillies à la fin de  $k$  intervalles fixés  $(0, t_1), \dots, (t_{k-1}, t_k)$ . Du fait que l'observation d'ex-aequos parmi les temps de censure et de survie résulte d'un regroupement des données, l'ordre réel de ces temps à l'intérieur d'un intervalle n'est pas connu ; la vraisemblance que proposent alors les auteurs est la probabilité que, sous le modèle (1), le vecteur des rangs de tous les temps de survie soit l'un de ceux qui sont compatibles avec les données comme elles sont observées.

Le Log de cette vraisemblance, qui ne dépend que de  $\beta$ , est d'une écriture assez lourde :

$$L_4(\beta) = \sum_{j=1}^k \beta' S_j + \sum_{j=1}^k \text{Log} \left[ \sum_{(p_1, \dots, p_{m_j}) \in Q_j} \left\{ \prod_{r=1}^{m_j} \frac{1}{\sum_{h \in R_j(p_r)} \exp(\beta' z_h)} \right\} \right]$$

où,  $Q_j$  est l'ensemble des permutations des  $m_j$  sujets morts à  $t_j$ , les sujets de l'une de ces permutations étant désignés par  $p_1, p_2, \dots, p_{m_j}$  ; d'autre part  $R_j(p_r)$  est l'ensemble des sujets à risque à  $t_j$ , "moins" les sujets  $p_1, p_2, \dots, p_{r-1}$ .

Il est possible, comme pour les méthodes précédentes, de calculer itérativement un estimateur  $\hat{\beta}$  de  $\beta$  maximisant  $L_4(\beta)$ .

### 3. METHODES UTILISABLES EN PRESENCE DE NOMBREUX EX-AEQUOS

#### 3.1. Discussion des méthodes I, II, III, IV

En l'absence d'ex-aequos parmi les temps de survie (soit, avec les notations du § 2 :  $m_j = 1$  pour tout  $j$ ) le calcul montre que les quatre Log de vraisemblance  $L_1, L_2, L_3$  et  $L_4$  sont égaux ; les quatre méthodes I, II, III, IV fournissent donc dans ce cas un même estimateur  $\hat{\beta}$  de  $\beta$ .

Un exemple de données de survie comprenant un nombre modéré d'ex-aequos à chaque temps de survie a été analysé par BRESLOW [2] : les méthodes II, III et IV ont été appliquées à ces données ; les résultats ne montraient que peu de différence entre les méthodes, tant en ce qui concerne les estimations  $\hat{\beta}$  que les statistiques de test de  $H_0 : (\beta = 0)$ .

En présence de nombreux ex-aequos, il apparaît que :

- les méthodes III et IV sont inapplicables ; en effet le calcul de  $n_j ! / m_j ! (n_j - m_j)!$  termes à chaque temps de survie  $t_j$  est nécessaire pour évaluer  $L_3(\beta)$  et ses dérivées ; de même, le calcul de  $m_j !$  terme à chaque  $t_j$  doit être effectué pour évaluer  $L_4(\beta)$  et ses dérivées ; les temps de calcul nécessités par chacune de ces deux méthodes deviennent prohibitifs quand les quantités  $m_j$  sont grandes.

- la méthode II n'est, en théorie, pas adaptée à cette situation ; en effet, dans cette méthode, les variables aléatoires  $T_{z_i}$  sont supposées continues, et la probabilité d'observer des ex-aequos est nulle.

Cependant, cette méthode, appliquée à l'exemple du § 4, conduit, comme on le verra, à des résultats très voisins de ceux obtenus avec les méthodes décrites en 3.2 et 3.3 ci-dessous.

### 3.2. Deux autres méthodes

Les deux méthodes décrites maintenant utilisent la vraisemblance calculée de la façon suivante :

les  $t_j$ , ( $j = 1 \dots k$ ) sont les extrémités de  $k$  intervalles donnés ; alors la probabilité que le sujet "i" mesure dans l'intervalle  $(t_{j-1}, t_j)$  est :

$$\{1 - p_1(z_i)\} \dots \{1 - p_{j-1}(z_i)\} p_j(z_i)$$

où

$$p_j(z_i) = P \{T_{z_i} \in (t_{j-1}, t_j) \mid T_{z_i} > t_{j-1}\}$$

D'autre part, la probabilité que le temps de survie du sujet "i" soit supérieur au temps  $c$  ou il est censuré vaut, pour  $c \in (t_{j-1}, t_j)$  :

$$\{1 - p_1(z_i)\} \dots \{1 - p_{j-1}(z_i)\} \{S(c, z_i)/S(t_{j-1}, z_i)\}$$

où  $S(t, z_i)$  est la fonction de survie de  $T_{z_i}$  en  $t$ .

Il s'ensuit que la vraisemblance des données s'écrit :

$$V = \prod_{j=1}^k \left[ \prod_{i \in M_j} \{p_j(z_i)\} \prod_{i \in R_j - M_j - C_j} \{1 - p_j(z_i)\} \prod_{i \in C_j} \{S(c, z_i)/S(t_{j-1}, z_i)\} \right]$$

où  $M_j$  (resp.  $C_j$ ) désigne l'ensemble des sujets morts (resp. censurés) dans l'intervalle  $(t_{j-1}, t_j)$  et  $R_j$  désigne l'ensemble des sujets à risque au début du jème intervalle  $(t_{j-1}, t_j)$ . Cette vraisemblance peut prendre différentes formes selon l'équation définissant  $p_j(z_i)$  en fonction de  $z_i$  ; deux telles équations déterminent les méthodes V et VI. D'autre part dans ces deux méthodes deux hypothèses différentes sont faites concernant les temps de censure.

#### Méthode V

PRENTICE et GLOECKLER [10] considèrent le cas où les variables aléatoires  $T_{z_i}$  sont continues, satisfaisant (1), et où les données sont groupées par intervalles (même situation que dans la méthode IV).

De plus, les temps de censure, inconnus, compris dans chaque intervalle  $(t_{j-1}, t_j)$  sont considérés comme égaux à  $t_{j-1}$ .

Alors, la vraisemblance (4) devient :

$$V = \prod_{j=1}^k \left[ \prod_{i \in M_j} \{p_j(z_i)\} \prod_{i \in R_j - M_j} \{1 - p_j(z_i)\} \right] \quad (5)$$

sous le modèle (1),  $p_j(z_i)$  peut s'écrire :

$$p_j(z_i) = 1 - \exp \left\{ - \exp(\beta' z_i) \int_{t_{j-1}}^{t_j} \lambda_0(u) du \right\}$$

ou encore :

$$1 - p_j(z_i) = (1 - p_j)^{\exp(\beta' z_i)},$$

avec :  $p_j = p_j(0)$ .

En posant alors :

$$\text{Log}(1 - p_j) = - \exp \alpha_j$$

le log de la vraisemblance s'écrit :

$$L_5(\beta, \alpha_1, \dots, \alpha_k) = \sum_j \sum_{i \in M_j} \text{Log} [1 - \exp \{- \exp(\alpha_j + \beta' z_i)\}] - \sum_j \sum_{i \in R_j - M_j} \exp(\alpha_j + \beta' z_i)$$

Des estimateurs de  $\beta$  et des  $\alpha_j$  ( $j = 1, 2, \dots, k$ ) peuvent être calculés de façon itérative comme dans les méthodes précédentes, et le test de  $H_0 : (\beta = 0)$  peut être obtenu de la façon habituelle.

Dans le cas où aucun temps de survie n'est observé dans l'intervalle  $(t_{j-1}, t_j)$ , le calcul de la dérivée première de  $L_5$  par rapport à  $\alpha_j$  montre que  $\alpha_j$  est estimé par  $-\infty$ ; tous les termes concernant cet intervalle s'annulent, de sorte qu'un tel intervalle est ignoré.

## Méthode VI

THOMPSON Jr. [11] a proposé un modèle où  $p_j(z_i)$  s'écrit sous forme logistique en fonction de  $z_i$ , comme dans la méthode III; il est plus simple, pour la suite des calculs, d'écrire la relation (3) sous la forme :

$$p_j(z_i) = \exp(\alpha_j + \beta' z_i) / \{1 + \exp(\alpha_j + \beta' z_i)\} \quad (6)$$

avec

$$\alpha_j = \text{Log} \{p_j / (1 - p_j)\}$$

Les temps de censure sont pris en compte de la façon suivante :

— la fonction r.i.  $\lambda(t, z_i)$  d'un sujet censuré à  $c$  avec  $c \in (t_{j-1}, t_j)$  est considérée comme constante sur l'intervalle  $(t_{j-1}, t_j)$ ; soit  $\lambda_j$  cette constante.

Alors,

$$\frac{S(c, z_i)}{S(t_{j-1}, z_i)} = \exp \{- \lambda_j (c - t_{j-1})\} = \left\{ \frac{S(t_j, z_i)}{S(t_{j-1}, z_i)} \right\}^{(c - t_{j-1}) / (t_j - t_{j-1})} \quad (7)$$

De plus, les temps de censure à l'intérieur de chaque intervalle sont considérés comme uniformément distribués sur cet intervalle; l'espérance du temps de censure d'un sujet censuré entre  $t_{j-1}$  et  $t_j$  vaut donc :

$$t_{j-1} + (t_j - t_{j-1})/2$$

THOMPSON remplace  $c$  par cette espérance dans (7); alors, et en écrivant  $a_{ji} = \alpha_j + \beta' z_i$ , le Log de la vraisemblance (4) vaut :

$$L_6(\beta, \alpha_1, \dots, \alpha_k) = \sum_{j=1}^k \left[ \sum_{i \in M_j} a_{ji} - \sum_{i \in R_j - C_j} \text{Log} \{1 + \exp(a_{ji})\} - \frac{1}{2} \sum_{i \in C_j} \text{Log} \{1 + \exp(a_{ji})\} \right]$$



Les paramètres  $\beta$  et  $\alpha_j$  ( $j = 1, 2, \dots, k$ ) sont estimés et le test de  $H_0 : (\beta = 0)$  est obtenu de la façon habituelle.

Dans le cas où aucun temps de survie n'est observé dans l'intervalle  $(t_{j-1}, t_j)$ ,  $\alpha_j$  est estimé par  $-\infty$ , et, comme dans la méthode V, un tel intervalle est ignoré.

### 3.3. Une simplification de la méthode VI

#### Méthode VII

Dans le cas où les données sont recueillies à la fin d'intervalle fixés, comme en 3.2, nous proposons une simplification de la méthode VI qui consiste, comme dans la méthode V, à supposer que, pour tout  $j$ , les temps de censure compris dans le jème intervalle sont égaux à  $t_{j-1}$ .

La vraisemblance est alors donnée par (5) où  $p_j(z_i)$  s'écrit sous la forme logistique (6). En posant, comme dans la méthode VI,  $a_{ji} = (\alpha_j + \beta'z_i)$ , le Log de la vraisemblance s'écrit :

$$L_7(\beta, \alpha_1, \dots, \alpha_k) = \sum_{j=1}^k \left[ \sum_{i \in M_j} a_{ji} - \sum_{i \in R_j} \text{Log} \{1 + \exp(a_{ji})\} \right]$$

Des estimateurs pour  $\beta$  et  $\alpha_j$  ( $j = 1, \dots, k$ ) sont obtenus itérativement de la façon habituelle. Pour tester  $H_0 : (\beta = 0)$  dans le cas particulier de la comparaison de deux groupes ( $z_i = 0$  ou 1 selon que le sujet "i" appartient à un groupe ou à l'autre), la statistique (2) devient :

$$V(\hat{\gamma})' W(\hat{\gamma}) V(\hat{\gamma}) = \left( M_1 - \sum_{j=1}^k m_j \frac{n_{1j}}{n_j} \right)^2 / \sum_{j=1}^k \left\{ m_j \left( \frac{n_j - m_j}{n_j} \right) \frac{n_{0j}}{n_j} \frac{n_{1j}}{n_j} \right\}$$

où  $\hat{\gamma} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_k, 0)$  est l'estimateur de  $\gamma = (\alpha_1, \alpha_2, \dots, \alpha_k, \beta)$  sous  $H_0$ , et  $M_1$  est le nombre total de morts dans le groupe "1",  $n_{1j}$  (resp.  $n_{0j}$ ) est le nombre de sujets à risque au début du jème intervalle dans le groupe "1" (resp. dans le groupe "0"),  $n_j$  et  $m_j$  ayant le même sens que précédemment.

La statistique que nous avons établie ci-dessus est très proche de celle de MANTEL, où le terme  $(n_j - m_j)/n_j$  au dénominateur est remplacé par  $(n_j - m_j)/(n_j - 1)$  [9].

## 4. UN EXEMPLE

Les données considérés concernent 308 sujets opérés pour valvulopathie aortique [1]; les temps de survie et de censure ont été enregistrés mois par mois à partir de la date d'opération.

Les covariables retenues sont : l'âge au moment de l'opération, le rapport cardio-thoracique ( $\times 100$ ), et la présence ou l'absence d'extrasystoles ventriculaires juste avant l'opération.

Les temps de survie distincts observés s'échelonnent entre 1 et 119 mois et sont au nombre de 31; 29 sujets sont morts pendant le premier mois; les nombres d'ex-aequo aux temps postérieurs à 1 mois sont inférieurs ou égaux à 4.

Les méthodes III et IV, demandant un temps de calcul trop long du fait de la présence des 29 ex-aequo à 1 mois n'ont pas pu être utilisées.

Les méthodes II, V, VI et VII ont été appliquées aux données ; les estimateurs des paramètres  $\beta_1$ ,  $\beta_2$ , et  $\beta_3$  correspondant aux trois covariables — dans cet ordre —, leur écart-type asymptotique ainsi que les rapports normaux de ces deux quantités sont donnés au Tableau I.

Il apparait que les méthodes V (Prentice — Gloeckler), VI (Thompson Jr.) et VII (logistique non conditionnelle) mènent à des résultats très voisins ; les résultats donnés par la méthode II (Breslow) sont légèrement différents.

La fonction de survie d'un sujet de rapport cardio-thoracique et d'âge moyens, et sans extra systole ventriculaire, a été estimée par les quatre méthodes ; les résultats, non reproduits ici, sont très voisins pour les quatre méthodes.

TABLEAU I  
Estimation des paramètres  
(les écarts-types sont entre parenthèses après chaque valeur estimée,  
et les rapports normaux sont sous les estimateurs).

| <i>Méthode</i>       | $\beta_1$            | $\beta_2$            | $\beta_3$            |
|----------------------|----------------------|----------------------|----------------------|
| II Breslow           | 0.037(0.014)<br>2.63 | 0.048(0.018)<br>2.60 | 0.595(0.296)<br>1.99 |
| V Prentice-Gloeckler | 0.038(0.014)<br>2.75 | 0.052(0.018)<br>2.78 | 0.604(0.301)<br>2.00 |
| VI Thompson Jr.      | 0.039(0.014)<br>2.72 | 0.052(0.019)<br>2.68 | 0.651(0.314)<br>2.07 |
| VII Logistique       | 0.039(0.014)<br>2.72 | 0.053(0.019)<br>2.70 | 0.649(0.314)<br>2.07 |

## 5. CONCLUSION

Les trois méthodes V, VI et VII sont théoriquement et pratiquement appropriées pour analyser des données de survie en présence de nombreux ex-aequo ; les temps de calcul diffèrent peu d'une méthode à l'autre, et les résultats obtenus sur notre exemple ne montrent que de très légères différences.

La méthode VII est, conceptuellement la plus simple des trois, tout en présentant l'avantage d'être la plus facile à programmer.

D'autre part, il faut remarquer aussi que les résultats obtenus par ces trois méthodes ne diffèrent qu'assez peu des résultats obtenus avec la méthode de BRESLOW (Méthode II), bien que celle-ci suppose une distribution continue pour les temps de survie.

## BIBLIOGRAPHIE

- [1] J. ACAR, P. LUXEREAU et P. DUCIMETIERE. — Les sténoses aortiques opérées : appréciation du pronostic à distance de l'intervention en fonction des paramètres cliniques et hémodynamiques. *Arch. Mal. Cœur*, 72, 973-982, 1979.
- [2] N. BRESLOW. — Applications to clinical trials of multiple regression models for censored survival data. *Communication présentée. "The Scientific Meeting of the Committee on Controlled Therapeutic Trials of the U.I.C.C."* April 1973.
- [3] N. BRESLOW. — Covariance analysis of censored survival data. *Biometrics*, 30, 89-99, 1974.
- [4] N. BRESLOW. — Analysis of survival data under proportionnal hazards model. *Int. Stat. Review*, 43, 45-88, 1975.
- [5] D.R. COX. — Regression models and life tables. *J.R. Statist. Soc., B*, 34, 187-220, 1972.
- [6] D.R. COX. — Partial likelihood. *Biometrika*, 62, 269-276, 1975.
- [7] J.D. KALBFLEISCH and R.L. PRENTICE. — Marginal likelihoods based on Cox's regression and life model. *Biometrika*, 60, 267-278, 1973.
- [8] R. KAY. — Proportionnal hazard regression models and the analysis of censored survival data. *Appl. Statist.*, 26, 227-237, 1977.
- [9] N. MANTEL. — Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 50, 163-170, 1966.
- [10] R.L. PRENTICE and L.A. GLOECKLER. — Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34, 57-67, 1978.
- [11] W.A. THOMPSON Jr. — On the treatment of grouped observations in life studies, *Biometrics*, 33, 463-470, 1977.